Numerical accuracy of Ox 6.2

Kulaksizoglu, Tamer

19 April 2012

Numerical Accuracy of Ox 6.21

Tamer Kulaksizoglu*

April 19, 2012

Abstract

This article evaluates the numerical accuracy of Ox object-oriented matrix programming language. It uses the Standard Reference Datasets (StRD) for estimation accuracy, the ELV DOS program for statistical distribution accuracy, and the DIEHARD battery of randomness tests for random number generation accuracy. The main finding of the paper is that Ox is quite accurate in almost every area it is tested. However, it should also be stressed that there is room for improvement in its optimization library.

1 Introduction

Ox is an object-oriented matrix programming language developed by Dr. Jürgen Doornik and designed after the popular languages C, C++, and Java. The main advantages of Ox can be listed as follows: it is free for academic research and teaching, it has rich mathematics, statistics, optimization, and graphics libraries, it is fast, its syntax is well-designed, and it is object-oriented. The formal sources to learn more about Ox are [3], [4], and [5]. Though, as a scientific programming language, Ox has been around since 1990s and reviewed by [1], [6], and [2], to the best of our knowledge, its numerical accuracy has never been subject of any study. This paper aims to fill that gap.

2 Ox's Numerical Accuracy

We evaluate Ox's numerical accuracy on three frontiers. The first frontier involves the well-known Statistical Reference Datasets (StRD). These data sets serve as reference data sets with certified

*The Ox code, the data sets, and the computer output used in this work are available from the author upon request at tamer.kulaksizoglu@gmail.com.

Table 1: Univariate Summary Statistics

Data Set	Difficulty	Mean	Standard Deviation	Autocorrelation
PiDigits	Lower	15.00	15.00	15.00
Lottery	Lower	15.00	15.00	14.94
Lew	Lower	15.00	15.00	14.84
Mavro	Lower	15.00	13.12	13.94
Michelso	Lower	15.00	13.85	13.44
NumAcc1	Lower	15.00	15.00	15.00
NumAcc2	Average	15.00	15.00	15.00
NumAcc3	Average	15.00	9.46	12.23
NumAcc4	Higher	15.00	8.25	11.03

results for the purpose of objective evaluation of statistical estimation routines[1]. The StRD contains five groups of data sets, namely, summary statistics, analysis of variance, Markov chain Monte Carlo, linear regression, and nonlinear regression. Following [9], we measure numerical accuracy with the logarithm of relative error (LRE):

$$LRE = \begin{cases} -log_{10}\left(|e-c|/|c|\right), & \text{if } c \neq 0 \\ -log_{10}\left(|e|\right), & \text{if } c = 0 \end{cases}$$

where e is the estimated value obtained from the program and c is the correct value. The second frontier aims to evaluate Ox's numerical reliability of statistical distributions. Here we use [7]'s ELV DOS program. Finally, the third frontier evaluates Ox's random number generators using [8]'s DIEHARD tests.

2.1 Univariate Summary Statistics

The univariate summary statistics data group contains nine data sets with the number of observations ranging from 3 to 5,000. Six of these data sets are of lower difficulty, two are of average difficulty, and one is of higher difficulty. For each data set, certified values are provided for mean, standard deviation, and first-order autocorrelation coefficient. For the mean, we use the function `meanc`, which takes a matrix as an argument and computes the column averages. For the standard deviation, we use the function `varc`, which takes a matrix as an argument and computes the column variances. We compute the standard deviation by multiplying the function return with $n/\left(n-1\right)$, where n is the number of observations, and then taking the square root. For the first-order autocorrelation, we use

[1]The data sets and the certified values are available at `http://www.itl.nist.gov/div898/strd/general/dataarchive.html`.

Table 2: Analysis of Variance

Data Set	Difficulty	F-Statistic
SiRstv	Lower	12.94
SmLs01	Lower	14.99
SmLs02	Lower	14.95
SmLs03	Lower	14.99
AtmWtAg	Average	11.70
SmLs04	Average	9.29
SmLs05	Average	9.29
SmLs06	Average	9.29
SmLs07	Higher	3.27
SmLs08	Higher	3.27
SmLs09	Higher	3.27

the function `acf`, which takes a matrix and an integer as arguments. The integer signifies the order of the autocorrelation, which is set to 1 in the calculations.

Table 1 shows the number of correct significant digits. As can be seen from the table, Ox's numerical accuracy for the mean is perfect, scoring 15 out of 15 for each data set. The standard deviation computations are also quite good. Ox has a little bit difficulty with only two data sets, namely, NumAcc3 and NumAcc4. As to the first-order autocorrelation calculations, Ox's numerical performance is totally satisfactory, again showing only a minor weakness for the aforementioned data sets. Overall, the results show that Ox is quite accurate for the univariate summary statistics.

2.2 Analysis of Variance

The analysis of variance data group contains eleven data sets with the number of observations ranging from 25 to 18,009. Four of these data sets are of lower difficulty, four are of average difficulty, and three are of higher difficulty. Each data set is balanced with only one treatment variable. For each data set, certified values are provided for sum of squares, degrees of freedom, mean squares, F-statistic, R-squared, and residual standard deviation for both within- and between-treatment. Following [9], we present the results only for the F-statistic. However, before presenting the results, we should note that Ox does not have a built-in analysis of variance function. For that reason, we create a class named `ANOVA` to perform the calculations, which uses only the built-in functions, based on the formulas provided at the URL `http://itl.nist.gov/div898/strd/anova/SiRstv_cmd.html`.

Table 2 shows the the number of correct significant digits. For the data sets with lower level of difficulty, except for SiRstv, Ox's numerical accuracy is almost perfect. For the data sets with average

Table 3: Markov Chain Monte Carlo

Data Sets	mcmc01	mcmc02	mcmc03	mcmc04	mcmc05	mcmc06
Difficulty	Lower	Lower	Average	Average	Higher	Higher
$E(\mu\|\overline{y}, s)$	15.00	15.00	15.00	15.00	15.00	15.00
$\sqrt{Var(\mu\|\overline{y}, s)}$	7.83	6.45	5.24	4.82	3.44	2.23
$q_{0.025}(\mu)$	15.00	15.00	15.00	15.00	15.00	15.00
$q_{0.5}(\mu)$	15.00	15.00	15.00	15.00	15.00	15.00
$q_{0.975}(\mu)$	15.00	15.00	15.00	15.00	15.00	15.00
$E(\sigma\|\overline{y}, s)$	7.83	6.45	5.24	4.82	3.44	2.23
$\sqrt{Var(\sigma\|\overline{y}, s)}$	7.83	6.45	5.24	4.82	3.44	2.23
$q_{0.025}(\sigma)$	7.83	6.45	5.24	4.82	3.44	2.23
$q_{0.5}(\sigma)$	7.83	6.45	5.24	4.82	3.44	2.23
$q_{0.975}(\sigma)$	7.83	6.45	5.24	4.82	3.44	2.23

level of difficulty, there is a visible decrease in accuracy. And for the data sets with higher level of difficulty, the accuracy gets even lower. A careful examination of the data sets reveals the following conclusion: the higher the data values in a data set are, the lower the accuracy of Ox gets. However, Ox's overall accuracy for the analysis of variance is still acceptable.

2.3 Markov Chain Monte Carlo

The Markov Chain Monte Carlo data group contains six data sets, each of which has 11 observations. Two of these data sets are of lower difficulty, two are of average difficulty, and two are of higher difficulty. For each data set, certified values are provided for posterior mean, posterior standard deviation, and 2.5%, 50%, and 97.5% posterior quantiles of the population mean μ and standard deviation σ. We create a class named `MCMC` to calculate these values based on the formulas provided at the URL `http://itl.nist.gov/div898/strd/mcmc/mcmc01_cmd.html`. The class uses several built-in functions, namely, the gamma function `gammafact`, the Chi-square and the t quantile functions `quanchi` and `quant` as well as the functions to take the square, square root, and means.

Table 3 shows the the number of correct significant digits. Unlike the other tables, this table shows the data sets in columns and the statistics in rows since there are less data sets than there are statistics in this data group. $E(\bullet|\overline{y}, s)$ represents the posterior mean, $\sqrt{Var(\bullet|\overline{y}, s)}$ the posterior standard deviation, and $q_\alpha(\bullet)$ the αth posterior quantile where $\bullet$ stands for either μ or σ.

The results show a clear pattern. For all the statistics, except for the posterior mean and quantiles of μ, the higher the level of difficulty of a data set is, the lower the degree of Ox's accuracy gets. Again, the loss of accuracy can be attributed to the fact that the data values are bigger for the data sets with higher level of difficulty. For instance, while the integer part of the data values in `mcmc01` is

Table 4: Linear Regression

Data Set	Difficulty	Coefficient		Standard Error	
		QR	Choleski	QR	Choleski
Norris	Lower	13.10	12.25	13.95	14.39
Pontius	Lower	-	11.11	-	13.24
NoInt1	Average	14.72	14.72	15.00	15.00
NoInt2	Average	15.00	15.00	15.00	14.78
Filip	Higher	-	-	-	-
Longley	Higher	12.01	7.32	13.49	8.45
Wampler1	Higher	9.84	7.31	10.36	7.85
Wampler2	Higher	13.86	11.82	15.00	13.17
Wampler3	Higher	9.86	7.31	14.27	10.72
Wampler4	Higher	9.07	7.31	14.30	10.72
Wampler5	Higher	7.00	7.31	14.31	10.72

only 9 digits, it is 14 digits in `mcmc06`. As for the the posterior mean and quantiles of μ, Ox obtains the perfect score for all the data sets.

2.4 Linear Regression

The linear regression data group contains eleven data sets with the number of observations ranging from 3 to 82 and the number of parameters ranging from 1 to 11. Two of these data sets are of lower difficulty, two are of average difficulty, and seven are of higher difficulty. For each data set, certified values are provided for coefficients, their standard errors, residual standard deviation, R-squared, and the analysis of variance table. Following [9], we present the minimum values of the LREs for the coefficients and the standard errors.

To carry out the actual calculations, we create a class named `OLS`. The class uses the built-in ordinary least squares functions `ols2c` and `olsc`, which solves the normal equations with the Choleski and the QR decompositions, respectively. In Ox's documentation, it is stated that `olsc` is numerically more stable, which seems to be the case.

Table 4 shows the the number of correct significant digits. Ox's numerical accuracy is quite satisfactory even for the data sets with higher level of difficulty. However, the QR decomposition fails for the Pontius data set, which is of lower level of difficulty. Also both decompositions fail for the Filip data set. A possible explanation is that, for both data sets, the explanatory variables contain very large values. For instance, the maximum value is 9×10^{12} for Pontius and it is $2,726,901,792.45$ for Filip. For Pontius, the return value from estimation is -2, which indicates that the issue is a combination of "ratio of diagonal elements of $\mathbf{X}'\mathbf{X}$ is large", with an error code of 2, and "$\mathbf{X}'\mathbf{X}$ is (numerically) singular", with an error code of -1. For the error code 2, Ox advises "rescaling" the data, which does work. Dividing the explanatory variables, excluding the intercept term, by 10 produces LREs of 13.95 and 13.54 for the coefficients and the standard errors, respectively. For Filip, the return value from estimation is -2 for the QR decomposition and 2 for the Choleski decomposition. However, no rescaling makes the problem go away, which may be a sign that Ox perceives $\mathbf{X}'\mathbf{X}$ as numerically singular.

Overall, Ox's numerical reliability for the linear regression gets high marks. However, the readers are advised to check the return value from their estimation and make sure that it is 1, which indicates success. If it is not, rescaling the data may be considered.

2.5 Nonlinear Regression

The nonlinear regression data group contains twenty seven data sets with the number of observations ranging from 6 to 250 and the number of parameters ranging from 2 to 9. Eight of these data sets are of lower difficulty, eleven are of average difficulty, and eight are of higher difficulty. For each data set, there are three sets of starting values, two of which are not close to the certified values and the third one is the certified values. The numerous nature of the data group and the difficulty of figuring out the analytical first derivatives make this group the hardest to code. To simplify the task, we use the analytical derivatives reported by [11]. To minimize the objective function, we make use of the BFGS algorithm implemented in the built-in `MaxBFGS` function. Ox also has the Newton-Raphson algorithm, implemented in the built-in `MaxNewton` function, but this algorithm requires the second-order derivatives and gives decomposition failure error messages when used only with the first-order derivatives. Besides, based on our trials on a few data sets, it does not seem to produce more accurate results than the BFGS algorithm.

Before presenting the results, we would like to offer a few words of advice for the reader on coding the objective function and its gradient to estimate nonlinear least squares in Ox. It seems that Ox's BFGS algorithm is a bit sensitive to mathematical expressions. Based on our experience, minor changes in the code may lead to different results. Having said that, we think it would be a good idea not to create long expressions, split them into smaller segments, and then combine them in a single variable if necessary. It would also be useful to try different, and possibly simpler, versions of the same mathematical expression. For instance, for an expression like $b_2x + x^2$, the reader should also try the mathematically equivalent $(b_2 + x)\,x$ and see if they produce different results.

Table 5: Nonlinear Regression

Data Set	Difficulty	Numerical Derivatives			Analytical Derivatives		
		Start 1	Start 2	Start 3	Start 1	Start 2	Start 3
Misra1a	Lower	2.77	2.63	15.00	11.65	11.17	15.00
Chwirut2	Lower	5.54	5.54	14.91	9.49	10.19	15.00
Chwirut1	Lower	5.84	5.86	10.82	10.76	9.27	15.00
Lanczos3	Lower	0.32^b	0.68^a	15.00	0.54^b	0.50^b	15.00
Gauss1	Lower	7.31	7.31	13.81	11.11	10.70	10.97
Gauss2	Lower	7.22	7.22	12.90	10.90	10.92	14.50
DanWood	Lower	$-^b$	6.50	15.00	$-^b$	6.26	15.00
Misra1b	Lower	3.24	2.05	15.00	10.87	11.21	15.00
Kirby2	Average	0.71^b	0.99^b	7.73	0.71^b	0.99^b	15.00
Hahn1	Average	$-^b$	$-^b$	6.95	$-^b$	$-^b$	15.00
Nelson	Average	5.15	1.46^b	15.00	10.86	1.46^b	15.00
MGH17	Average	0.34^a	3.96	13.73	0.34^a	4.18	15.00
Lanczos1	Average	0.32^b	0.90^a	15.00	0.56^b	0.49^b	15.00
Lanczos2	Average	0.33^b	0.90^a	15.00	0.56^b	0.49^b	15.00
Gauss3	Average	7.25	7.25	15.00	10.73	10.59	14.82
Misra1c	Average	1.51	1.61	15.00	10.71	11.11	15.00
Misra1d	Average	2.81	2.23	15.00	10.47	11.13	15.00
Roszman1	Average	6.61	6.73	15.00	6.31	1.92	15.00
ENSO	Average	8.95	9.07	15.00	9.03	10.09	15.00
MGH09	Higher	$-^b$	4.11	15.00	$-^a$	3.87	15.00
Thurber	Higher	6.84	$-^b$	12.83	10.58	$-^b$	13.49
BoxBOD	Higher	9.90	0.71^b	15.00	10.50	0.71^b	15.00
Rat42	Higher	$-^b$	7.50	15.00	15.00	10.37	15.00
MGH10	Higher	$-^b$	$-^b$	11.29	$-^b$	$-^b$	10.82
Eckerle4	Higher	0.97^b	7.51	15.00	0.97^b	8.21	15.00
Rat43	Higher	$-^b$	9.60	10.11	$-^b$	10.86	15.00
Bennett5	Higher	1.78^b	1.16^b	11.79	0.68^b	0.39^b	15.00

$^{a(b)}$Modifying the tolerance level does (not) improve the results.

Table 5 shows the number of correct significant digits for the parameters with the default tolerance level, which is 1E-04. Once again, the reported digits belong to the least accurate estimates. The character "-" signifies that the accuracy is zero or negative. For the seven data sets with lower level of difficulty, Ox has issues with only the two, namely, Lanczos and DanWood[2]. However, for the data sets with average level of difficulty, Ox's accuracy declines significantly. In that category, Ox has issues with six of the eleven data sets. Finally, Ox has issues with all of the eight data sets in the category with higher level of difficulty. As can be seen from the table, only a few of the low-accuray results can be improved by modifying the tolerance levels. By way of comparison, [10] reports more accurate LREs for SAS, SPSS, and S-Plus[3]. These results suggest that Ox's numerical optimization library has room for improvement.

2.6 Statistical Distributions

Another area where numerical accuracy is important is statistical distributions. Ox has a very rich library of distribution functions. To evaluate them numerically, we use [7]'s ELV DOS program to generate correct digits[4].

Table 6 shows the exact and estimated values for the central and non-central F, non-central Chi-square, and non-central Student's t distributions[5]. Our findings show that the central and non-central F distributions seem to be correct for probabilities as small as 1E-14 and 1E-12, respectively. Similarly, the non-central Student's t seems to be accurate up to 14 digits. Finally, the non-central Chi-square seems to be correct up to 21 digits but suddenly returns zero beyond that point. Overall, Ox's numerical reliability for statistical distributions is quite good.

2.7 Random Number Generation

Finally, we evaluate Ox's random number generators (RNG). For that purpose, we use the DIEHARD battery of randomness tests[6]. [10] explains the testing process in detail. Ox implements five random number generators. Namely, they are

[2] The issue with this data set goes away if the objective function is written as `y - exp(log(b1) + b2*log(x))` instead of the more straightforward `y - b1 * pow(x, b2)`.

[3] We should emphasize that SAS and S-Plus use the Gauss-Newton algorithm whereas SPSS uses the Levenberg-Marquardt algorithm.

[4] The DOS executable and the documentation are available at the URL `http://www.statistik.lmu.de/~knuesel/elv/index.htm`.

[5] In our tests, Ox had no problems with the standard normal, Gamma, Chi-square, Beta, Student's t, Poisson, Binomial, and Hypergeometric distributions so we do not report their results.

[6] The URL is `http://stat.fsu.edu/pub/diehard/`.

Table 6: Statistical Distributions

$F(2,100)$				Non-central $F(2,100)$			
x		Exact	Estimated	x	λ	Exact	Estimated
30		0.622302E-10	0.622301E-10	30	199	0.109436E-7	Exact
31		0.334387E-10	Exact	31	200	0.196585E-7	0.196586E-7
32		0.181054E-10	Exact	30	200	0.917210E-8	0.917211E-8
33		0.987633E-11	0.987632E-11	31	204	0.983756E-8	0.983757E-8
34		0.542670E-11	0.542666E-11	30	213	0.887399E-9	0.887418E-9
35		0.300299E-11	0.300304E-11	31	217	0.988409E-9	0.988418E-9
Non-central $t(3)$				Non-central $\chi^2(20)$			
x	δ	Exact	Estimated	x	λ	Exact	Estimated
5	20	0.236257E-8	Exact	9	90	0.119184E-15	Exact
6	24	0.120059E-8	Exact	10	91	0.423178E-15	Exact
5	21	0.275226E-9	0.275225E-9	9	91	0.805652E-16	0
6	25	0.189612E-9	0.189610E-9	10	95	0.908310E-16	0
5	22	0.287477E-10	0.287464E-10	9	97	0.762707E-17	0
6	26	0.276896E-10	0.276884E-10	10	101	0.892703E-17	0

- LCG31: Modified Park and Miller with period 2^{31}
- MWC60: George Marsaglia with period 2^{60}
- LFSR113: Pierre L'Ecuyer with period 2^{113}
- MWC8222: George Marsaglia multiply-with-carry with period 2^{8222}
- MWC8222_52: George Marsaglia multiply-with-carry with period 2^{8222}

The default RNG is MWC8222_52. The test results are shown in Table 7.

Ox's random number generators pass all of the 18 tests with one exception: the Modified Park and Miller RNG fails the Binary Rank for 32×32 Matrices test. Based on these results, we can conclude that Ox's RNGs, especially MWC8222 and MWC8222_52, are quite suitable for studies involving heavy-duty simulations.

3 Conclusion

In this paper, we evaluate Ox's numerical accuracy on three frontiers: statistical estimation, statistical distribution, and random number generation. The results suggest that Ox's numerical accuracy is quite satisfactory, which makes it the perfect environment for econometric research. The only area where Ox is not so impressive is nonlinear least squares estimation. We should stress, however, that this may not be a serious issue. The BFGS algorithm implemented in Ox is a general optimization

Table 7: Random Number Generation

Test	PM	GM	LE	MWC_32	MWC_52
Birthday Spacings Test	P	P	P	P	P
Overlapping 5-Permutation Test	P	P	P	P	P
Binary Rank for 31×31 Matrices	P	P	P	P	P
Binary Rank for 32×32 Matrices	**F**	P	P	P	P
Binary Rank for 6×8 Matrices	P	P	P	P	P
Bitstream Test (p values)	P	P	P	P	P
OPSO Test	P	P	P	P	P
OQSO Test	P	P	P	P	P
DNA Test	P	P	P	P	P
Count-the-Ones Test (stream of bytes)	P	P	P	P	P
Count-the-Ones Test (specic byte)	P	P	P	P	P
Parking Lot Test	P	P	P	P	P
Minimum Distance Test	P	P	P	P	P
3-D Spheres Test	P	P	P	P	P
Squeeze Test	P	P	P	P	P
Overlapping Sums Test	P	P	P	P	P
Runs Test	P	P	P	P	P
Craps Test	P	P	P	P	P

algorithm. More dedicated algorithms for nonlinear least squares estimation such as Gauss-Newton and Levenberg-Marquardt may produce more accurate results and implementing them in Ox should not be a challenging task.

References

[1] F. Cribari-Neto. Econometric programming environments: GAUSS, Ox, and S-Plus. *Journal of Applied Econometrics*, 12:77–89, 1997.

[2] J. A. Doornik. Object-oriented programming using Ox. In S. S. Nielsen, editor, *Programming Languages and Systems in Computational Economics and Finance*, volume 18 of *Advances in Computational Economics*, chapter 5, pages 115–148. Kluwer Academic Publishers, 2002.

[3] J. A. Doornik. *An Object-Oriented Matrix Language: Ox 6*. Timberlake Consultants Press, 2009.

[4] J. A. Doornik and M. Ooms. *Introduction to Ox*, 2006.

[5] J. A. Doornik and M. Ooms. *Introduction to Ox: An Object-Oriented Matrix Language*. Timberlake Consultants Press, 2007.

[6] T. Kenc and J. M. Orszag. Ox: An object-oriented matrix language. *The Economic Journal*, 107:256–259, 1997.

[7] L. Knüsel. *Computergestützte Berechnung Statisticher Verteilungen*. Oldenbourg, München, 1989.

[8] G. Marsaglia. DIEHARD: A battery of tests of randomness. `http://www.stat.fsu.edu/pub/diehard/`, 1996.

[9] B. D. McCullough. Assessing the reliability of statistical software: Part I. *The American Statistician*, 52(4):358–366, 1998.

[10] B. D. McCullough. Assessing the reliability of statistical software: Part II. *The American Statistician*, 53(2):149–159, 1999.

[11] B. D. McCullough. Econometric software reliability: EViews, LIMDEP, SHAZAM, and TSP. *Journal of Applied Econometrics*, 14:191–202, 1999.